

Efficient Embedding of Information and Knowledge into CSCL Applications

Santi Caballé¹, Thanasis Daradoumis¹ and Fatos Xhafa²

¹Open University of Catalonia, Department of Computer Science and Multimedia
Rbla Poblenou, 156. 08018 Barcelona, Spain
{scaballe, adaradoumis}@uoc.edu

²Department of Languages and Informatic Systems
Polytechnic University of Catalonia,
Jordi Girona Salgado 1-3, 08034 Barcelona, Spain
fatos@lsi.upc.es

Abstract. This study aims to explore two crucial aspects of collaborative work and learning: the importance of enabling CSCL applications, on the one hand, to capture and structure the information generated by group activity and, on the other hand, to extract the relevant knowledge in order to provide learners and tutors with efficient awareness and support as regards collaboration. To this end, we first identify and define the main types of information generated in on-line group activity and then propose a process for efficiently embedding this information and the knowledge extracted from it into CSCL applications for awareness and feedback purposes. The conceptual model proposed finally gave rise to the design and implementation of a CSCL generic platform, called the Collaborative Learning Purpose Library (CLPL), which serves as a basis for the systematic development of collaborative learning applications and for providing full support to the mentioned process of knowledge management.

1 Introduction

Computer-Supported Collaborative Learning (CSCL) is an emerging paradigm [1] for research in educational technology that focuses on the use of information and communications technology (ICT) as a mediation tool within collaborative methods of learning. When developing CSCL environments that support online collaborative learning, several issues must be taken into account in order to ensure full support to the online learning activity. One such key issue is interaction data analysis, a core function for the support of coaching and evaluation in CSCL environments. It relies on information captured from the actions performed by participants during the collaborative process [2]. The efficient embedding of this information and of the extracted knowledge into CSCL applications sets the basis for enhancing monitoring, awareness [3] and feedback [4] to achieve a successful learning process in collaborative environments.

Therefore, the success of CSCL applications depends to a great extent on the capability of such applications to embed information and knowledge of group activity and use it to achieve a more effective group monitoring. In the literature, however, ques-

tions related to the information and knowledge embedding have not been sufficiently investigated. An initial approach [4] considers the use of feedback in on-line learning and its effects on group activity in general. To this end, some types of information generated by the group activity are considered as relevant knowledge to be communicated to the group members for feedback purposes but the process of how to collect the information, analyze it and extract the desired knowledge is not provided.

CSCL applications are characterized by a high degree of user-user and user-system interaction and hence are generating a huge amount of event information. This information can be appropriately collected and automatically processed by computers as a data source to extract relevant knowledge of the collaboration. Note that in this context information refers to the event data generated by the learning group and knowledge refers to the result of the treatment of this information in terms of analysis techniques and interpretations that will be presented to the same group that generated it.

Therefore, the event information management is the cornerstone in this context, aiming at facing three main problems: (i) how to obtain and classify the necessary information gathered from the collaborative activity into several description levels; (ii) how to efficiently process the large amount of information collected during the group activity in order to facilitate its later analysis and make the extracted knowledge available to the participants even in real time; (iii) how information should be analyzed and what kind of knowledge should be extracted to be fed back to the participants in order to provide the best possible support and monitoring of their learning and instructional processes. Finally, we need to provide an efficient and robust computational approach that enables the embedding of the collected information and the extracted knowledge into a CSCL application.

In order to achieve these goals, we first propose a conceptual model for data analysis and management that identifies and classifies the many kinds of indicators (variables) that describe collaboration and learning into several high-level potential categories of effective collaboration. Then, we enter a process that, as a first step, collects and classifies the event information generated by the group activity according to these indicators. For efficiency purposes, this information may then be structured in a way that facilitates its later processing and analysis. The last stage of this process consists of interpreting the analysis outcomes and communicating the knowledge extracted to the group members for awareness and feedback purposes as well as to the tutors to effectively track the collaborative learning process.

Achieving a clear and well-structured conceptual model can greatly facilitate the design of effective computational models that reflect as accurately as possible task performance, individual and group behavior, interaction dynamics, members' relationships and group support. As a consequence, a generic platform for the systematic construction of CSCL applications is provided. This platform, called the Collaborative Learning Purpose Library (CLPL) [5], translates the conceptual model and its indicators into a computational model of collaborative learning interaction and can be used to embed information and knowledge into CSCL applications in an efficient manner.

The paper is organized as follows. We first present a conceptual model that identifies and classifies the three main categories of information generated in group activity. In Sect. 3 we show how the online information collected and classified can go through a process of analysis with the aim of extracting all the necessary knowledge that is to

be presented to the interested actors. This process and the categories defined in the conceptual model are translated into a computational model based on a CSCL generic platform presented in Sect. 4. We conclude in Sect. 5 with some comments and point out ongoing and further work.

2 A Conceptual Model for Managing Group Activity Information

The conceptual model we propose shows how the information generated in collaborative learning activities can be captured and classified at several description levels. This fact can significantly improve the way a collaborative system used for learning and instruction can collect all the necessary information produced from the user-user and user-system interaction in an efficient manner. The aim is to model different aspects of interaction and thus at helping all the actors involved understand the outcomes of the collaborative process.

To this end, we classify group activity information into three generic categories of activity: the members' contributing behavior to the task (the *outcome of collaboration*), the *functioning of the group* (the interaction processes underlying the collaborative learning activities, such as participation behavior, role playing, etc.), and *individual and group scaffolding* (social support and task- or group functioning-oriented help). We summarize these issues in this section (for more details, see [6]).

Collaborative Learning Outcome (or Task Performance). This category represents the mid- and low-level indicators in the form of the skills and sub-skills that should characterize the students who participate in a learning collaborative situation in order to achieve effective group and individual performance of the task and thus obtain a successful learning outcome. To measure each indicator (or skill), we associate it with the students' actions, which represent each indicator in the best possible manner. We employ a similar terminology to the one usually used in the BSCW system [7] to refer to the actions that can be carried out in any groupware platform. Even so, they are general enough to be independent from BSCW and represent all the typical and basic actions encountered in every groupware platform.

Group Functioning. This category includes the mid- and low-level indicators in the form of skills and sub-skills that students should exhibit in order to enhance participation, promote better communication and coordination, and achieve the effective interaction and functioning of the group in a learning collaborative situation. To measure each indicator, we associate it with specific student actions which best describe each skill to be accomplished.

Scaffolding. This category includes the different types of social support and help services [8] that have been identified and accounted for in our model. The participants' actions aiming at getting or providing help are classified and measured according to whether they refer to the task or group functioning.

3 The Process of Embedding Information and Knowledge into a CSCL Application

In this section, we present a process to provide a learning group with relevant knowledge extracted from learners' interaction data in CSCL applications for awareness, feedback and monitoring purposes. The aim is to greatly improve the effectiveness of the learning exercise.

Here, we face two difficult problems: First, the problem of how to define an efficient process of embedding information and knowledge into a computer-mediated collaboration given that several essential steps need taking into account. Second, how to give relevant and semantically grounded feedback on what is happening in a collaborative learning framework to students and teachers in order to allow them eventually to modify the on-going activity. We focus on and discuss a solution to the first problem, while providing some suggestions as how to deal with the second.

The process of correctly providing information and knowledge in collaborative learning applications involves three separate, necessary steps: collection of information, analysis and presentation (see Fig. 1 and [9]). The entire process fails if any one of these steps is omitted. During the first step, a structuring and classification of the generated event information is needed. This information is then analyzed in order to extract the desired knowledge. The final step is to provide users with the essential awareness and feedback from the obtained knowledge.

3.1 The Collection and Classification of Event Information.

The most important issue while monitoring group activity is the collection and storage of a large amount of event information generated by the high degree of interaction among the group participants. Such a large amount of informational data may need a long time to be processed. Therefore, collaborative learning systems have to be designed in a way that classifies and pre-structures the resulting information in order, on the one hand, to correctly collect the group activity and, on the other hand, to increase the efficiency during data processing in terms of analysis techniques and interpretations. Due to its importance, the processing step is treated in great detail later.

As shown in the previous section classification of the information is achieved by distinguishing several high-, mid- and low-level indicators of effective collaboration. Based on this, we further categorize and specify users' particular actions according to the following criteria:

- Who is doing something? (i.e. the originator of the event).
- When did s/he do it? (i.e. timestamp).
- Where does s/he do it? (i.e. the location of the affected resource).
- What is s/he doing? (i.e. the action type and the object involved).
- Why did s/he do it? (i.e. student intentions and motivation which are captured by the indicators associated with each action; e.g., a user performs the action "create document" in order to generate new information in the shared workspace).

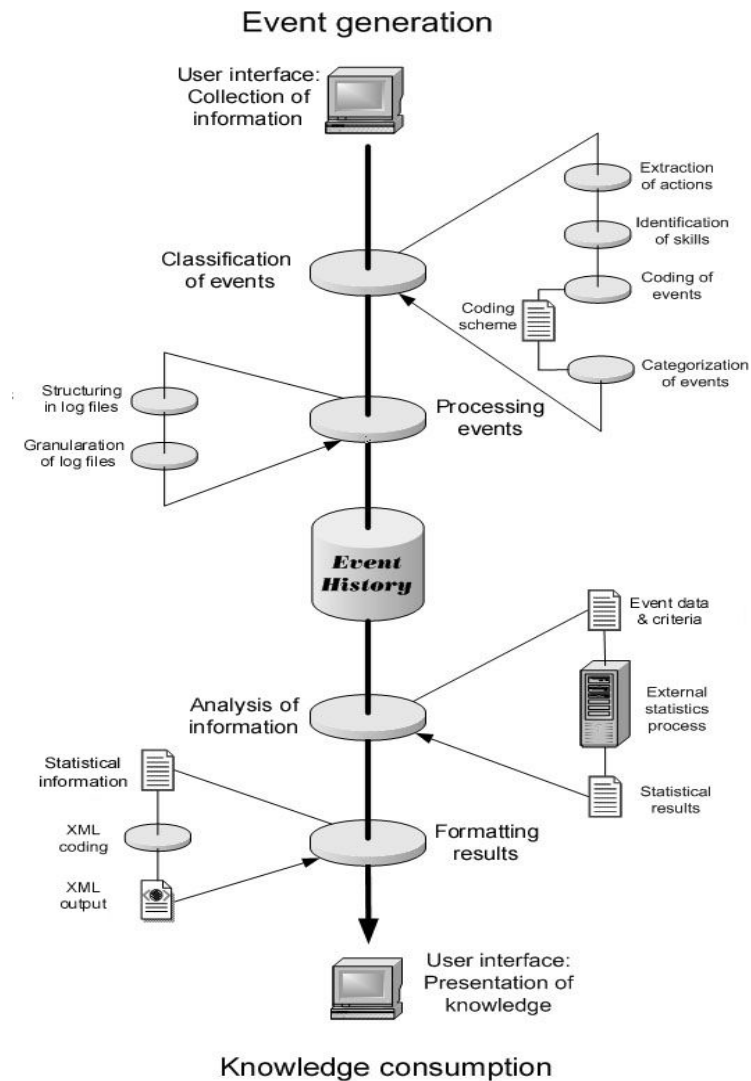


Fig. 1. The process of embedding information and knowledge into CSCL applications

Our aim is to provide a guideline to correctly classify the user actions on the resources during group activity. To this end, we propose a classification process in which the event information collected from the log files is handled in sequential steps consisting of *extraction*, *identification*, *coding*, and *categorization* (see Fig. 1). In particular, we first *extract* the specific action performed by a user on a resource (e.g. file document, debate, etc.). Second, this action is interpreted according to the type of event that was involved in (this information should be provided implicitly by the sys-

tem according to the context where the action was triggered or explicitly by the user who triggered it). This provides the basic information that is used for the *identification* of the real intentions or skills shown by the user (e.g. creating a contribution during a debate can be interpreted as either revision or reinforcement of the information depending on whether the contribution was created in the context of a reply or as an observation). Subsequently, we *codify* the user event taking into account both the user action and the event type. Doing so, we associate a unique code to the user skill identified in the context of the action. Finally, we *categorize* the user event into one of the group activity indicators defined in Sect. 2 and shown in Fig. 2.

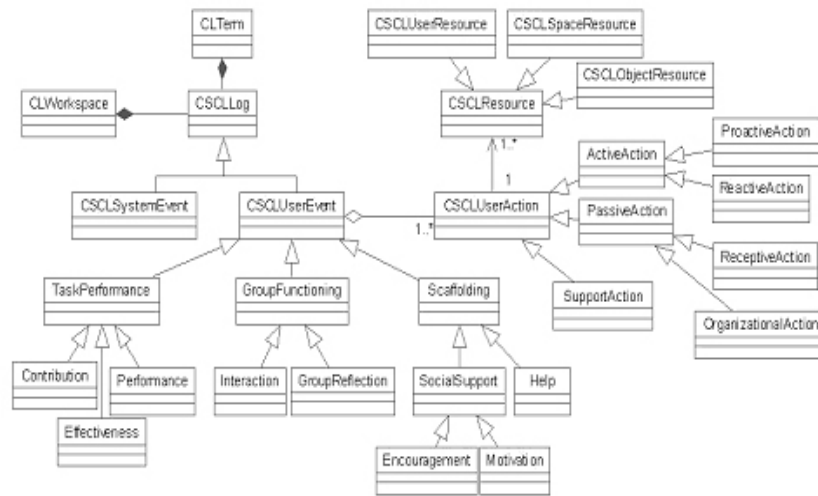


Figure 2. A hierarchy to collect and classify all events generated during the group activity.

3.1.1 Efficient Processing of the Information

Due to the large amount of event information generated in CSCL applications, once this information activity has been correctly collected and classified we may come across the issue of demanding computational requirements while processing this information. In order to facilitate this step, CSCL applications may structure this information as log files in a way that takes advantage of the parallelism in order to process several files (e.g. all the groups in a classroom) at the same time and thus dramatically reduce the overall computational time to process them [10], [11]. As a result, it is possible for these applications to process a large volume of collaboration activity data and make the extracted information available even in real time. Note that this step is optional within this process of embedding information and knowledge and it is proposed for efficiency purposes only.

To this end, we propose the following generic steps so as to correctly structure the event information for later processing (see Fig. 1 and [10]): by classifying the event information and turning it into persistent data, we store it in the system as structured files. These files contain all the information previously collected from the system log files in specified fields. Next, we predefine the structured files in accordance with certain criteria such as time and workspace, which characterize all group collaboration. Our goal is to achieve a high degree of granularity of log files. Thus, during later data processing, it is possible to concatenate several structured files so as to obtain the appropriate degree of granularity (e.g. all groups in a classroom for each 12 hours). This makes it possible to efficiently parallelize data processing depending on the characteristics of the computational resources.

To resume, the process of embedding information and knowledge into collaborative learning applications came to the point where the information generated by group activity has been collected, classified and well-structured so that it can be easily and efficiently processed and analyzed during the second stage of this process. We now turn to present this stage whose purpose is to extract relevant knowledge to be fed back to the participants.

3.2 Data Analysis and Extraction of Knowledge.

The second stage of this process consists in processing all the information previously collected and classified according to the indicators mentioned before by means of analysis techniques. There is a fair deal of research focused on the analysis of online group interaction. A detailed description of an integrated approach we followed to analyze the collaborative learning interaction can be found in [11]. As a consequence of this analysis, knowledge is generated providing meta-cognition about the state and evolution of interaction, which enhances awareness about the efficiency of group activity, group behavior and the individual attitudes of its members in shared workspaces.

Knowledge extraction is based on criteria related to the three socio-cognitive functions that operate simultaneously during group interaction, namely *production function*, *group well-being* and *member support* [12] and their associated indicators. In that sense, as regards the production function, we can extract knowledge by constantly observing the members' activities (e.g. showing each group member's absolute and relative amount of contributions) or the status of shared resources. In addition, we can obtain knowledge that is relevant to individual and group well-being by exploring the communication and interaction flow among group members (such as members' motivational and emotional state, comparative studies of effective and ineffective groups and so on). Finally, knowledge can be acquired by ill-functioning situations, such as missing or insufficient contributions, lack of participation, etc., which can reveal the need for helping individual members by providing them specific scaffolding where and when this is necessary (i.e. member support).

The definition of a variety of indicators at several levels of description allows us to determine the granularity of information to be transmitted to the interested parties. In

other words, based on a model of desired interaction (establishing a comparison of the current and desired state of interaction), the analysis approach detects and highlights the indicators which were not satisfied and need to be corrected by redirecting group and individual attitudes. These indicators reveal those aspects of the collaborative learning activity (task performance, group functioning, or scaffolding) that present problems and need to be corrected adequately. Thus, they set up rules and filters in order to extract and summarize only that information which refers to the malfunctioning aspect. The summarized information is finally transformed into useful knowledge that is communicated to and acquired by the group members who use it to improve the performance of the problematic aspect.

Therefore, on the one hand, our approach enables group members to become aware of the progress of their peers in performing the learning exercise both at individual and group level, as well as of the extent to which other members are participating in the collaborative process as this influences their *decision making* [13]. On the other hand, our approach provides tutors with information about students' problem-solving behavior, group processing [14] and performance analysis [15] for assessment and guiding purposes [16]. This approach is presented below and constitutes the last stage of the process of embedding information and knowledge into CSCL applications.

3.3 Presentation of the Knowledge Acquired.

Here the problem consists in identifying the roles and needs of each learner and the tutor in every moment and being able to decide what information is required to be provided, in which granularity and how to present it. For example, the knowledge obtained from the interaction analysis should be tailored in such a way that the support provided for self-regulation or peer assessment is adapted to the role the learner plays at a particular moment. In that way, scaffolding information would be different for a learner playing a coordinator role from one that plays a software designer role. Moreover, the format used to present the information could vary from case to case.

Consequently, we proceed to define three different levels that dictate how the acquired knowledge is to be presented, that is, at what format and detail level:

- **Awareness level.** At this level, we need to inform participants about what is going on in their shared workspace, providing information about their own actions or the actions of their peers, or presenting a view of the group interaction, behavior and performance [3]. To this end we display plane indicator values that show the state and specific aspect of the collaborative learning interaction and processes that take place. The information presented to the learner can support him/her at a meta-cognitive level.
- **Assessment level.** At this level, we need to provide data and elements to assess the collaborative activity, so the indicators used are associated with specific weights that measure the significance of each indicator in the assessment process. As in the previous case, the information provided acts at a meta-cognitive level, giving the actors the possibility to evaluate their own actions and behavior as well

as the performance of their peers and the group as a whole.

- **Scaffolding (or Guiding) level.** Supporting participants during collaborative activities has become a main concern of current research [4], [17]. At this level, we need to produce information aiming at guiding, orienting and supporting students in their activity. This information is determined by the unsatisfied indicators and helps students to diagnose problematic situations and self-estimate the appropriateness of their participation in a collaborative activity as well as to counsel their peers whenever insufficient collaboration is detected.

In this section we have discussed the conceptual process that defines how the embedding of information and knowledge can take place in a CSCL application in an efficient manner and how important it is to consider this knowledge as a means to providing awareness, feedback as well as individual and group monitoring. We now turn to discuss the implementation of this approach by a computational model based on a generic platform that is used both for the systematical construction of CSCL applications and for embedding information and knowledge from group activity into them. This platform is presented in the next section.

4 A Computational Model to Embed Information and Knowledge

We propose a generic, robust, reusable component-based *Collaborative Learning Purpose Library*¹ (CLPL) [5], [18] based on the Generic Programming paradigm [19] so as to enable a complete and effective reutilization of its generic components for the construction of specific CSCL applications. We use this platform as a computational model especially for the implementation of the conceptual model of information management and the process of embedding this information and the knowledge extracted into CSCL applications as described in Sect. 2 and 3.

The CLPL consists of generic components related to user management, administration, security, knowledge management, and functionality mapping the essential needs in which any CSCL application is involved. In this context, we are especially interested in using this platform as a computational model for data analysis and management. This is mainly performed by two components, namely *CSCL Knowledge Management* and *CSCL Functionality* components, which form the core of the CLPL in the construction of CSCL applications. Due to its importance, they are described here briefly (see [5] for details).

¹ CLPL is found at: <http://cv.uoc.edu/~scaballe/clpl/api/index.html>

4.1 CSCL Knowledge Management Component

In developing our CLPL we paid special attention to event analysis and management. To this end, this component is made up of two subsystems, namely *CSCL Activity Management* and *CSCL Knowledge Processing* so as to support the first two stages of the process of embedding of information and knowledge described in Sect. 3. The third stage, presentation (of user awareness and feedback), is accomplished by a different component called *CSCL Functionality* that we will describe in the next subsection.

The ultimate objective of this component is to define a bottom-up analysis approach that analyses the user events in order to decode the specific actions of the users describing their interaction during the collaboration activities. The analysis aims at identifying those sequences of actions that can be used to determine typical patterns of interactions [20].

At this point, based on a model of desired interaction, the system allows us to compare the learners' real interaction processes with the typical interaction patterns in order to infer whether or not the process is effective for the learner. Furthermore, the knowledge revealed by this analysis can enhance self and peer evaluation, which in turn improves the efficiency of group activities, monitoring group behavior and the individual attitudes of its members in the shared workspace. Finally, this knowledge is also essential in assisting the tutor by providing the necessary means to support and assess individual and group learning outcomes.

Next component provides the suitable means to present the results of this analysis so that all the actors of the collaboration can consume and benefit from the knowledge acquired.

4.2 CSCL Functionality Component

This component, which has five subsystems in all, defines the three elemental parts involved in any form of cooperation, namely *coordination*, *communication* and *collaboration* [18]. *Coordination* involves the organization of groups to accomplish the important objectives of members such as workspace organization and group structure and planning. *Collaboration* lets group members share any kind of resources while *communication* represents the basis of the whole component since it enables coordination and collaboration to be achieved by providing them with low-level communication support.

The final objective of this component is to provide functional support to CSCL applications in terms of group organization, resource sharing, user interaction, and so on. Moreover, this component implements the last stage of the process of embedding information and knowledge into CSCL applications by presenting the knowledge generated to users in terms of immediate awareness and constant feedback of what is going on in the system.

5 Conclusions and Future Work

In this paper we have shown the importance of providing efficient knowledge and information from collaborative learning activity and how to efficiently embed them into CSCL applications as crucial aspects in the enhancement of the collaborative learning process. To this end, we have first described a conceptual model that captures and classifies three main types of information generated in the group activity by means of potential indicators of effective collaboration and secondly we presented the process of embedding this information and the knowledge extracted into CSCL applications. Then, the categories defined in the conceptual model are translated into a computational model of collaborative learning interaction by means of a generic platform called *Collaborative Learning Purpose Library* for the construction of collaborative learning applications which is specially used to embed information and knowledge into CSCL applications in an efficient manner. The CLPL library has been evaluated by using it as a basis to develop a *Web-based structured discussion forum* [5] to support on-line collaborative learning activities. We plan to use this application in our real context of the Open University of Catalonia to gain experience with it.

Further work will focus on how to incorporate information retrieval and document filtering techniques into the stage of information collection to automatically extract knowledge from information with a high degree of informality (e.g. to evaluate the group's well-being function) and thus making it possible to embed it into CSCL applications. We would also like to investigate how to integrate a portable, general and reusable CSCL ontology in our generic platform as a declarative representation of the knowledge embedded into CSCL applications with the aim to both describe how these systems are built and understand how real groups work.

Acknowledgements. This work has been partially supported by the Spanish MCYT project TSI2005-08225-C07-05.

References

1. Koschmann, T: Paradigm shifts and instructional technology: An introduction. In T. Koschmann (Ed.) CSCL: Theory and practice of an emerging paradigm, 1-23. Mahwah, NJ: Lawrence Erlbaum Associates (1996)
2. Dillenbourg, P.: Introduction; What do you mean by "Collaborative Learning"? In P. Dillenbourg (Ed.), Collaborative learning. Cognitive and computational approaches (p. 1-19). Oxford: Elsevier Science (1999)
3. Gutwin, C., Stark, G. and Greenberg, S.: Support for Workspace Awareness in Educational Groupware. in Proceedings of the ACM Conference on Computer Supported Collaborative Learning, Bloomington, Indiana, USA October 17-20, 1995.
4. Zumbach, J., Hillers, A. & Reimann, P. (2003). Supporting Distributed Problem-Based Learning: The Use of Feedback in Online Learning. In T. Roberts (Ed.), Online Collaborative Learning: Theory and Practice pp. 86-103. Hershey, PA: Idea.

5. Caballé, S., Daradoumis, Th., and Xhafa, F.: A Model for the Efficient Representation and Management of Online Collaborative Learning Interactions. In P. Cunningham and M. Cunningham (Eds.), *Building the Knowledge Economy: Issues, Applications and Case Studies*. IOS Press, Amsterdam, Netherlands. (2006) 1485-1492
6. Daradoumis, T., Martínez, A. & Xhafa, F.: A Layered Framework for Evaluating Online Collaborative Learning Interactions. *International Journal of Human-Computer Studies*. Special Issue on "Theoretical and Empirical Advances in Groupware Research", Academic Press: Elsevier Ltd (2006)
7. R. Bentley, Hosrtmann, T., Trevor, J.: The World Wide Web as enabling technology for CSCW: The case of BSCW. In: *Computer-Supported Cooperative Work: Special Issue on CSCW and the Web*, vol. 6, 1997, 111-134
8. Webb, N.: Testing a theoretical model of student interaction and learning in small groups. In: R. Hertz-Lazarowitz and N. Miller (Eds.), *Interaction in Cooperative Groups: The Theoretical Anatomy of Group Learning*. Cambridge Univ. Press, NY (1992) 102-119.
9. Caballé, S., Daradoumis, T., Paniagua, C., & Xhafa, F.: A Grid Approach to Provide Effective Awareness to On-line Collaborative Learning Teams. *Proceedings of the First International Kaleidoscope Learning GRID Special Interest Group Workshop on Distributed e-Learning Environments*. Naples, Italy (2005)
10. Xhafa, F., Caballé, S., Daradoumis, Th. and Zhou, N.: A Grid-Based Approach for Processing Group Activity Log Files. In: *proc. of the GADA'04*, Agia Napa, Cyprus. LNCS 3292 (2004) 175-186
11. Caballé, S., Paniagua, C., Xhafa, F., & Daradoumis, Th.: A Grid-aware Implementation for Providing Effective Feedback to On-line Learning Groups. In: *proc. of the GADA'05*, Agia Napa, Cyprus. LNCS 3762 (2005) 274-284
12. Daradoumis, T., Martínez, A. & Xhafa, F.: An integrated approach for analyzing and assessing the performance of virtual learning groups. In: *Proc. of the 10th Int. Workshop on Groupware (CRIWG 2004)*, Berlin: Springer (2004)
13. McGrath, J.E.: Time, Interaction and Performance (TIP). *A Theory of Groups*. *Small Group Research*, 22, (1991) 147-174.
14. Dillenbourg, P. (ed.) (1999): *Collaborative Learning. Cognitive and Computational Approaches*. Elsevier Science Ltd. 1-19.
15. Kiesler, S. and Sproull, L.S. (Eds.): *Computing and change on campus*. New York: Cambridge Press (1987)
16. Daradoumis, T., Xhafa, F. and Marquès, J.M.: Exploring Interaction Behaviour and Performance of Online Collaborative Learning Teams, 9th Int. Workshop on Groupware, CRIWG'03, France. *Lecture Notes in Computer Science*, Vol. 2806, pp. 126-134 (2003)
17. Ellis, C., Gibbs, S., and Rein, G. Groupware: some issues and experiences. *Communications of the ACM*, 34,1, 9-28, 1991.
18. Lund, K. Human Support in CSCL : what, for whom, and by whom ? In J.-W. Strijbos, P. Kirscher & R. Martens (Eds.). *What we know about CSCL in Higher Education*. pp. 167 - 198. Dordrecht : Kluwer Academic Publishers, 2004.
19. Caballé S., Xhafa, F., Daradoumis, T. and Marquès, J.M. : Towards a Generic Platform for Developing CSCL Applications Using Grid Infrastructure. In: *Proc. of the CLAG/CCGRID'04*, Chicago, USA (2004)
20. Caballé, S. and Xhafa, F.: A Study into the Feasibility of Generic Programming for the Construction of Complex Software. 5th YRWS. In: *Proc. of the GPCE/NODE'03* pp.441-446, Erfurt, Germany (2003)
21. Inaba, A., Ikeda, M., & Mizoguchi, R.: What Learning Patterns are Effective for a Learner's Growth? - An ontological support for designing collaborative learning. In: *Proceedings of the International Conference on Artificial Intelligence in Education (AIED2003)*, pp. 219-226, Sydney, Australia, July 20-24 ,2003